

What is claimed is:

1. A method for determining an e-mail address formatting rule corresponding to a domain name, the method comprising:
 - 5 gathering e-mail address data corresponding to the domain name;
 - determining the e-mail address formatting rule based on the gathered e-mail address data;
 - and
 - electronically storing an association of the e-mail address formatting rule with the domain name.
- 10 2. The method of claim 1 wherein the step of gathering e-mail address data further comprises the steps of:
 - 15 providing a domain registration interface to a party having authority for the domain name, the domain registration interface including an interface for indicating one or more e-mail address formatting rules associated with the domain name; and
 - gathering the one or more e-mail address formatting rules associated with the domain name from the domain registration interface.
- 20 3. The method of claim 1 wherein the step of gathering e-mail address data further comprises the steps of:
 - registering e-mail addresses pursuant to an e-mail forwarding service; and
 - sorting registered e-mail addresses by domain name.
- 25 4. The method of claim 1 wherein the step of gathering e-mail address data further comprises the steps of:
 - accessing public e-mail address listings;
 - storing e-mail addresses from the public e-mail address listings.

5. The method of claim 1 wherein the step of gathering e-mail address data further comprises the steps of:

gathering e-mail addresses from one or more e-mail address books.

6. The method of claim 5 wherein the step of gathering e-mail address from one or more e-mail address books further comprises performing address correction on the one or more e-mail address books, and informing an address-book owner of any corrections to be made on the one or more e-mail address books.

7. The method of claim 1 wherein the e-mail address data comprises a first e-mail address, and the step of determining the e-mail address formatting rule further comprises:

parsing a domain portion of the first e-mail address;

parsing an identifier portion of the first e-mail address; and

determining whether the identifier portion is consistent with one or more known e-mail address formatting rules.

8. The method of claim 7 wherein the step for determining the e-mail address formatting rule further comprises:

for a plurality of e-mail addresses having the same domain portion as the first e-mail address, performing the steps of

parsing a plurality of identifier portions of the plurality of e-mail address;

determining whether the plurality of identifier portions are consistent with known e-mail address formatting rules; and

recording the known e-mail address formatting rules that are determined to be consistent with the plurality of identifier portions; and

counting a frequency at which particular known e-mail formatting rules were found to be consistent with the plurality of identifier portions; and

selecting from the recorded known e-mail address formatting rules, the e-mail address formatting rule for the domain name, based on the counted frequency.

9. The method of claim 7 wherein the step of determining whether the identifier portion is consistent with one or more known e-mail address formatting rules further comprises:

comparing the identifier portion to a list of known names;

identifying some or all the identifier portion as consistent with one or more known names;

determining whether the identifier portion is consistent with one or more known e-mail address formatting rules that are a function addressees' names.

10. The method of claim 9 wherein the step of determining whether the identifier portion is consistent with one or more known e-mail address formatting rules further comprises:

assigning first probabilities that some or all of the identifier portion corresponds to one or more names on the list of known names, based in part on a frequency of the names in a general population;

assigning second probabilities that one or more known e-mail address formatting rules are applicable to the first e-mail address, the second probability being a function of the first probability; and

selecting the e-mail address formatting rule for the first e-mail address based on the known e-mail address formatting rule that has the best second probability.

11. The method of claim 9 wherein the step for determining the e-mail address formatting rule further comprises:

for a plurality of e-mail addresses having the same domain portion as the first e-mail address, performing the steps of

parsing a plurality of identifier portions of the plurality of e-mail address;

comparing the plurality of identifier portions to the list of known names;

identifying some or all the plurality of identifier portions as consistent with one or more known names;

assigning first probabilities that some or all of the plurality of identifier portions corresponds to one or more names on the list of known names, based in part on a frequency of the names in a general population;

assigning second probabilities that one or more known e-mail address formatting rules are applicable to one of the plurality of e-mail addresses, the second probability being a function of the first probability; and

assigning third probabilities that one or more known e-mail address formatting rules are applicable to the plurality of e-mail addresses based on the cumulative second probabilities; and

wherein the step of selecting the e-mail address formatting rule for the domain of the first e-mail address is based on the known e-mail address formatting rule that has the best third probability.

12. The method of claim 9 wherein the list of known names includes first names and last names, and the step of comparing the identifier portion to a list of known names further comprises:

comparing a first sub-portion of the identifier portion to the list of known first names; and
comparing a second sub-portion of the identifier portion to the list of known last names.

13. The method of claim 12 wherein the first sub-portion is separated from the second sub-portion by a separator character.

14. The method of claim 7 wherein the identifier portion has a integer quantity of m characters, an integer n being smaller than or equal to m , and wherein the step of determining whether the identifier portion is consistent with one or more known e-mail address formatting rules further comprises:

(a) comparing the first n characters of the identifier portion to a list of known names to determine whether the first n characters of the identifier portion are consistent with the first n characters in one or more names on the list of known names;

(b) for a particular value of n , recording whether the first n characters of the identifier portion are consistent with the first n characters of the one or more names on the list of known names;

(c) changing the value of n , but not to be greater than m , and repeating steps (a) and (b);

(d) identifying values of n for which the identifier portion is consistent with the one or more names on the list of known names, including a maximum number of first characters for which the identifier portion is consistent with the one or more names on the list of known names;

(e) identifying whether the one or more names which were found to be consistent with the first n characters are first names or last names; and

(f) determining an identifier portion format as having a beginning character group being up to the maximum number of characters in a first or last name, as identified in step (e).

15. The method of claim 14 further comprising the steps of:

(g) performing steps (a) through (f) for a plurality of e-mail addresses having the same domain portion;

(h) determining a format rule for the domain, based on the frequency of identifier portion formats, including first or last name and the maximum numbers of characters, as determined in step (g).

16. The method of claim 14 wherein the step of determining whether the identifier portion is consistent with one or more known e-mail address formatting rules further comprises:

(g) comparing a remainder group of characters, after the beginning character group, to the list of known names;

(h) identifying whether the remainder group of characters is consistent with beginnings of one or more names from the list of known names;

(i) identifying a remainder maximum quantity of characters of the remainder group which were found to be consistent with the beginnings of one or more names from the list of known names

(j) identifying whether the remainder group of characters is consistent with first or last names from the list of known names; and

(k) determining the identifier portion format as having the remainder group having the maximum quantity of characters comprising the beginning letters of first or last names, as identified in step (j), the remainder quantity of characters positioned after the beginning character group.

17. The method of claim 16 wherein the step (k) of determining the identifier portion format further includes choosing the identifier portion format for the beginning character group and the remainder group such that they do not both correspond to a same name type, the same name type being last name or first name.

18. The method of claim 16 further comprising the steps of:

(l) performing steps (a) through (k) for a plurality of e-mail addresses having the same domain portion; and

(m) determining a format rule for the domain, based on the frequency of identifier portion formats, including those of beginning character groups and remainder groups for the plurality of e-mail addresses.

19. The method of claim 7 wherein the one or more known e-mail address formatting rules are from a list consisting of: last.first, first.last, alphanumeric only, LLLLLLFF, FFLLLLLL, FMLLLLLL, telephone number, punctuation required, minimum number of characters, or maximum number of characters.

20. The method of claim 1 wherein:

the step of gathering e-mail address data corresponding to the domain name includes gathering an e-mail address and addressee information about an addressee to whom a message is intended at the e-mail address; and

the step of determining the e-mail address formatting rule based on the gathered e-mail address data further includes comparing the addressee information to the e-mail address to derive the e-mail address formatting rule.

21. The method of claim 20 wherein the addressee information is a name of the addressee.

22. The method of claim 20 wherein the step of gathering e-mail address data includes gathering the addressee information from an address book.